The following memorandum is in response to Windward's summary of the Portland Harbor Benthic Predictive Approach meeting held on November 21, 2005 in Seattle. The comments provided here focus on general and specific technical issues identified by the government team. They are provided to 1) clarify for the LWG EPA's expectations with respect to various aspects of the benthic predictive approach and 2) to ensure that resources allocated to this effort are used in the most efficient manner practicable. In addition, EPA requests that LWG provide electronic copies of the presentations from the November meeting, including a summary of data manipulation and reduction efforts.

**Key Issues**:

*Floating Percentile Model Transparency*

- An important issue that must be addressed is that of transparency and replicability of the Floating Percentile Model (FPM). It is imperative that DEQ staff members are able to understand and replicate the results of the FPM. It has been suggested that appropriate DEQ and Windward staff work directly together to achieve the requisite transparency. EPA supports this approach and requests that arrangements be made as soon as possible to ensure that this occurs. Absent transparency, EPA will not be in a position to consider supporting the results of the FPM.

*Floating Percentile Model* Hyalella *Growth Endpoint*

- The *Hyalella* growth endpoint appears to be producing different results than other selected test endpoints. The LWG consultant has suggested that this endpoint be removed from the analysis based on the assertion that the FPM for this endpoint is not producing reliable results, even though it is being used as a part of the logistic regression modeling. EPA and its team members do not agree with this assessment. Poor model performance is insufficient reason to ignore the most sensitive endpoint. Furthermore, the pooled growth/survival results for the selected species are a useful way to look at the growth results because growth is not independent of survival (and mortality can be considered a lethal chronic endpoint). Therefore, the FPM should be applied to the pooled results for each species and it should address the combined endpoints for each species.

*Predicting Threshold Toxicity Values (e.g., PAHs)*

- The modeling results presented at the November meeting indicate that threshold values for PAHs are not playing a role in classifying samples. This interpretation could result in the establishment of threshold values that are well above ranges of concentrations associated with effects at other sites. EPA wants to clarify that we and our partners do not agree with LWG that these values indicate PAHs should be set at high threshold values. To resolve this issue, further investigation is needed in investigating other parameters that help explain the patterns in PAH toxicity (e.g., proportion of fines, inclusion of TPH fractions, etc.).

*N-qualified data*

- Some of the "N" qualified results are samples with high reported concentrations for several chemicals. These results should not be excluded without presenting supporting rationale.

*Floating Percentile Model Consideration of Hit/No-hit Thresholds and Consistency with Logistic Regression Model Thresholds*

- EPA would like to clarify that DEQ will run the FPM with hit/no-hit thresholds established at 90, 80, and 70 percent of controls to provide consistency with the logistic regression modeling (LRM) approach. EPA is not directing LWG to use these thresholds in the FPM but encourages consideration for doing so. Appropriate thresholds for the FPM will consider DEQ's analysis of EPA proposed hit/no-hit thresholds (70, 80, 90).

*Purpose of the Model: Predicting Toxicity vs. Predicting Non-toxicity*

- EPA would like to emphasize that the purpose of the model is to assist in remedial decision making by providing predictions of toxicity using sediment chemistry data. A model that does this effectively should provide, to the extent practicable, information about specific *contaminants*. A model that relies only on ammonia or sulfide or a limited group of contaminants (these, of course, can be used as indicators) may be of limited utility in developing clean up levels. If the model does not adequately predict toxicity, EPA may require additional bioassays to define areas of concern and establish appropriate cleanup levels.

**Technical Issues**

*Approach for and Application of Summing for Classes of Contaminant Data (e.g., PCBs, PAHs, DDTs, etc.)*

- LWG proposed in November to sum DDTs, PAHs and PCBs. Summing should be supported by the results of the modeling, and this should be presented clearly in the benthic approach report. However, EPA believes it would be better to sum at a later stage. For example, FPM analyses indicating that contaminants are correlated offers support for summation of contaminant classes. If contaminants are co-varying they will show this by where they "float" in the analysis. Strong correlations between individual PAHs and LPAH and HPAH could also support the use of LPAH and HPAH instead of individual PAHs. However, the report should present both the unsummed and summed analyses so the reader has the information needed to make a decision on the appropriateness of summing contaminant classes. The basis for summation should be described in the report, and summation decision rules should be consistent throughout the study area.

*Alpha Selection and Error Reduction*

- EPA notes that LWG is, in running its analyses, is using an alpha level of 0.05. The alpha levels represent the probability of making incorrect conclusions that the treated sample is toxic or contains chemical residues not found in the control or reference sample (Type 1 error). By setting this probability low (0.05), the likelihood that one erroneously concludes there are no differences among the mean responses in the treatment, control or reference samples (Type 2 error) increases (low power). Type 2 errors would lead to conclusions that the sample is not toxic (or different from control or reference), when in fact there is a difference. Type 2 errors are important to minimize in environmental investigations, since, if left undetected, these errors can lead to continued short- and long-term effects (ASTM 2003; EPA 2000a). In order to avoid this outcome and to be consistent with the workplan, an alpha of 0.1 should be used to increase the power of the test and the probability of detecting a reduction relative to the control mean.

*Methodology for Removing Contaminants from the Models*

- The LWG has proposed removing contaminants on the basis that they are not drivers of toxicity (e.g., aluminum). However, DEQ's analysis has shown that some were slight predictors of toxicity. The report and analysis should the complete list of detected contaminants and clearly justify the basis for removal from the modeling effort. The analysis (at least the FPM) should identify contaminants that are not useful predictors of toxicity throughout the study area, as exhibited by the calculated values, and this will provide justification for dropping contaminants

*Additional Lines of Evidence*

- The results of the bioassay tests and modeling effort have shown that additional lines of evidence may be important in interpreting the bioassay results (e.g., EqP or pore water testing).

*Control Normalized Results*

- EPA still does not agree with LWG’s proposed approach to control-normalization, which is to subtract the control result from the test result rather than dividing test by control as is commonly done. This approach is also being used for growth, which appears to be different than what was proposed in the original benthic methods memo. The impact on the results for this data set may be minimal, however it is EPA’s desire to create consistency between approaches (modeling efforts) as well as a defensible approach for determining relevant differences between control and test responses.